advances.in/psychology

August 30, 2024
Doi: 10.56296/aip00023
ISSN: 2976-937X

PERSPECTIVE

Causal inference in misinformation and conspiracy research

Li Qian Tay[1,2*], Mark Hurlstone[3], Yangxueqing Jiang[1], Michael J. Platow[1], Tim Kurz[2], & Ullrich K. H. Ecker[2,4]

Received: April 30, 2024 | **Accepted:** August 13, 2024 | **Published:** August 30, 2024 | **Edited by:** Jonas R. Kunst

[1] School of Medicine and Psychology, Australian National University; [2] School of Psychological Science, University of Western Australia; [3] Department of Psychology, Lancaster University; [4] Public Policy Institute, University of Western Australia. *Please address correspondence to Li Qian Tay, Li.Tay@anu.edu.au, School of Medicine and Psychology, The Australian National University, Canberra ACT 2600, Australia. This article is published under the Creative Commons BY 4.0 license. Users are allowed to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Psychological research has provided important insights into the processing of misinformation and conspiracy theories. Traditionally, this research has focused on randomized laboratory experiments and observational (non-experimental) studies seeking to establish causality via third-variable adjustment. However, laboratory experiments will always be constrained by feasibility and ethical considerations, and observational studies can often lead to unjustified causal conclusions or confused analysis goals. We argue that research in this field could therefore benefit from clearer thinking about causality and an expanded methodological toolset that includes natural experiments. Using both real and hypothetical examples, we offer an accessible introduction to the counterfactual framework of causality and highlight the potential of instrumental variable analysis, regression discontinuity design, difference-in-differences, and synthetic control for drawing causal inferences. We hope that such an approach to causality will contribute to greater integration amongst the various misinformation- and conspiracy- adjacent disciplines, thereby leading to more complete theories and better applied interventions.

Keywords: causal inference, conspiracy theory, fake news, methodological triangulation, misinformation

1. INTRODUCTION

Psychological research has revealed important insights into how individuals process misinformation and conspiracy theories—defined here as false or misleading information that runs counter to formal logic, objective evidence, or an established scientific consensus (see Ecker et al., 2024). Indeed, studies have identified an assortment of variables that are predictive of belief in misinformation and conspiracy theories, and a range of promising interventions have been proposed (Badrinathan & Chauchard, 2024; Douglas et al., 2019; Ecker et al., 2022; Newman et al., 2022; Kozyreva et al., 2023; Pennycook & Rand, 2022; Tay et al., 2023).

Nonetheless, the literature has focused mostly on laboratory-based experiments and observational (non-experimental) studies seeking to establish causality via third-variable adjustment, neglecting formal definitions of causality and the potential of drawing causal inferences from natural experiments (also see Tay et al., 2024). This is despite the fact that one of the earliest lessons scientists receive is that statistically significant associations do not necessarily imply causation (Grosz et al., 2020). The result is that, on the one hand, researchers in the field may sometimes be guilty of making (implied or otherwise) causal claims when they are not warranted; on the other hand, researchers may sometimes refrain from making causal claims—for example, in situations where randomized manipulation is impossible (e.g., due to ethics and feasibility considerations)—when there are in fact alternatives tools that researchers can use to test those causal claims even with non-experimental data (see Marinescu et al., 2018; Rohrer, 2018).

Considering that the testing of causal relations is critical in addressing a range of questions that researchers and practitioners alike may be interested in, including whether interventions against misinformation and conspiracy theories will improve subsequent individual or societal outcomes, it is our view that there is an urgent need for researchers in the field to draw from the broader literature on causal-inference methodologies. To this end, the current Perspective aims to provide a brief and accessible introduction to causal inference, as well as illustrate the potential of drawing causal inferences from natural experiments, using a mix of hypothetical and real-world examples relevant to misinformation and conspiracy research.

1.1 Why Should We Pay More Attention to Causal Inference?

First, the lack of attention to frameworks of causality has meant that there is, at present, no consensus on what can be deemed causal constructs and processes, nor sufficient reference points to anchor robust debates in the general literature. One salient example is the debate surrounding the framing of phenomena such as misinformation and conspiracy theories. In particular, some researchers have argued that misinformation is merely a "symptom" of societal conditions as opposed to a "cause" (e.g., Jungherr & Schroeder, 2021), while others have argued that misinformation has the potential to negatively impact (i.e., to causally affect) societal outcomes across a range of domains (e.g., Ecker et al., 2024; van der Linden et al., 2017). We believe that applying a formal definition of causality, namely a counterfactual definition, would allow researchers to be more precise in both the making and testing of the relevant claims.

Under the counterfactual definition, causality can be viewed as the outcome of comparisons between different states of the world. Here, an outcome Y can be said to be causally affected by event X, if, absent event X, outcome Y would have taken on a different value. This comparison between states can be formalized via the potential-outcomes framework (Rubin, 1974). For instance, to assess the causal effect of a piece of misinformation, we can contrast two potential outcomes for any individual, one for a world in which the individual is exposed to the misinformation and one in which the only difference is that there is no such misinformation.

If the individual is willing to pay $500 for an alternative-medicine intervention when exposed to misinformation but only $200 without exposure to misinformation, the misinformation caused a $300 increase in willingness-to-pay for the fictitious intervention; conversely, if outcomes do not differ, the misinformation would have had no effect.

Nonetheless, much of the research within misinformation and conspiracy literatures (and in psychology in general) has neglected to test causality, beyond the use of self-report measures and behavioral tasks within laboratory experiments (for recent reviews, see Goreis & Voracek, 2019; G. Murphy et al., 2023). This focus on laboratory experiments has placed unreasonable restrictions on the types of questions that we can credibly ask and answer, even though they may be of psychological and public interest (Grosz et al., 2020).

To advance misinformation and conspiracy research, we believe that it is important to disentangle the process of formal causal inference from that of experimentation. This is because formally defining causality in terms of counterfactual comparisons would allow researchers to make clearer the links between theory-implied causal relations and tested relations in our studies, and would allow us to better assess the plausibility of any underlying assumptions for when randomized manipulation is not possible. For instance, the adoption of a counterfactual definition highlights the "fundamental problem of causal inference," namely that an individual cannot simultaneously receive and not receive an intervention, and thus researchers often must find meaningful workarounds (see also Imbens & Rubin, 2015). One canonical theoretical quantity that researchers can target is therefore to estimate the average causal effect, which is the average difference in outcomes.[1] In practice, this can involve contrasting the average outcomes of a group of individuals that is exposed to the misinformation against a group of individuals that is not (i.e., a control group that represents the counterfactual). If individuals in the misinformation group are willing to pay an average of $500 for the spurious intervention, while individuals in the control group are willing to pay an average of $200, researchers can calculate the average causal effect of $300 even without individual-level data. Random assignment can render such an analysis unbiased by ensuring that an intervention is distributed independently with respect to potential outcomes, and that there will be comparable distribution of non-target characteristics across groups, so that any difference in average outcomes can be ascribed to the hypothesized cause. Randomized experiments are generally considered the "gold standard" for causal inference for this reason.

Yet, although randomized experiments represent the "gold standard", they may not always be possible. To illustrate, consider that a hypothetical team of researchers seek to study the impact of alternative-medicine misinformation targeting cancer patients. It is not possible for the researchers to conduct a randomized experiment that prescribes such misinformation, due to ethical concerns. The researchers may therefore instead conduct an experiment that tasks participants with reading a relevant vignette and imagining that they are a cancer patient assessing alternative treatment plans. The researchers may also randomly assign a subset of participants a counter-misinformation intervention (e.g., a debunking). To the extent that the study could be informative of potential interventions against cancer misinformation, one critical assumption is that the experimental paradigm is able to sufficiently represent the psychological processes involved, compared to if the participants actually had cancer. Such an assumption will not always be

[1] We note that targeting average effects in randomized experiments is one work-around; however, under the assumption that individuals can be used as their own controls, one could also devise plausible strategies to test for individual-level effects for at least some interventions.

tenable, as one can easily imagine that cancer patients may have different states of mind (see also hypothetical bias; Bernheim et al., 2022; J. Murphy et al., 2010). This limitation is independent of the fact that researchers can estimate the causal effect of the specific vignettes, and it applies whenever the real-world stakes and incentives (external or internal) exceed that which can be credibly manipulated by the experimenters. In fact, the most insidious form of impact may come not from isolated instances of exposure to misinformation and conspiracy theories, as they are typically implemented in time-constrained experimental studies, but rather from extended periods of influence with oft-repeated exposure from ostensibly trusted sources of media (see also Ash et al., 2024; Lewandowsky et al., 2017; Tay et al., 2024).

Critically, the specification of theoretical quantities as independent entities—for example, defining the aforementioned average causal effect as a counterfactual comparison between potential outcomes—would allow researchers to consider alternative means by which the quantities can be estimated, beyond randomized experiments. In this way, the effects that we seek to study can be better guided by logic, needs, and prior literature, as opposed to being constrained by any particular laboratory-based study design or empirical strategy (e.g., vignettes and regression modelling; Lundberg et al., 2021; MacCorquodale & Meehl, 1948). Indeed, study designs and empirical strategies then serve only as imperfect ways of estimating the theoretical quantities from data, and to what extent researchers should ultimately update beliefs about causal relations based on tests conducted on new data depends on whether the explicated assumptions are considered tenable.

The above has implications for the mutual-internal-validity problem faced by paradigms that focus only on laboratory experiments (Lin et al., 2021). The essence of this problem is that theories explaining only within-paradigm phenomena can gradually lose touch with meaningful outcomes beyond the paradigm if the same theories are always used to design the experiments that in turn guide development of the theories. Lin and colleagues argued that triangulation of methods (e.g., from self-report to behavioral and physiological) and theories (e.g., from psychology to economics and political science; see also Bor & Petersen, 2022) can be one way of addressing this problem (see also Haslam et al., 2020). In terms of methods, the use of a wider range of analysis and data sources can help address idiosyncratic artifacts arising from single sources (e.g., sample bias, measurement error, or context-specific influences); and in terms of theories, integrating insights from different disciplines can help refine existing models while inspiring new research (e.g., theories of evolution and continental drift were informed by disciplines as distinct as paleontology and geology). Clearer thinking about causality may, in our view, ensure that researchers are not unduly constrained to particularly empirical strategies and may thus help facilitate such triangulation in a systematic manner.

1.2 Additional Approaches for Causal Inference

Given the above, we now introduce several additional approaches of drawing causal inference that are currently underutilized, particularly within psychological research on misinformation and conspiracy theories. These strategies include instrumental-variable analysis, regression-discontinuity designs, as well as difference-in-differences and synthetic-control designs. There are research questions and data that may be more or less suitable for one strategy over others, depending on the causal structures of variables that researchers deem plausible and are willing to assume. Table 1 presents the two "standard approaches" in the field (i.e., randomized experiments and observational studies), alongside an overview of these additional strategies based around natural experiments. We also present a selection of relevant

Table 1

Selection of Empirical Strategies and Relevant Research

Strategy	Brief Explanation	Examples
Randomized experiment	- Randomization helps rule out alternative explanations and allows researchers to ascribe differences in outcomes to causal effects of the randomized manipulation. - Challenges arise when latent causes are hypothesized (e.g., psychological processes such as motivated reasoning; Tappin et al., 2020), and when ethical or feasibility considerations preclude certain manipulations (e.g., exposing cancer patients to cancer misinformation to test for misinformation effects).	- Effect of misinformation on vaccination intention (e.g., Loomba et al., 2021) - Effect of repetition on belief in true and false information (e.g., Pillai & Fazio, 2021) - Effect of source-credibility information and social norms on misinformation engagement (e.g., Prike et al., 2024)
Observational studies	- Observational studies rely on naturally occurring data or self-reports from research participants, and often seek to establish causality, either explicitly or implicitly, by controlling for third variables. - The choice of control variables should be explicitly justified, such that the assumption of no unmeasured confounding is plausible. However, it is instead often based on disciplinary norms and data availability, which can lead to confused analysis goals or unjustified causal conclusions for readers and the public, even if researchers avoid using causal terms (Grosz et al., 2020).	- Relationship between perceived (self-reported) exposure to misinformation and trust in institutions (e.g., Boulianne & Humprecht, 2023) - Relationship between social-media use and belief in conspiracy theories (e.g., Enders et al., 2021)
Instrumental-variable analysis	- Instrumental-variable analysis allows researchers to test for causal effect between an explanatory variable of interest and an outcome, even in the presence of unmeasured confounding between the two. - The analysis relies on identifying instruments, which are variables that influence the outcome only through their effects on the explanatory variable and that must not share	- Effect of media attention to terrorism on future terrorist attacks (e.g., Jetter, 2017) - Effect of watching Fox News during COVID-19 on social-distancing behaviors (e.g., Ash et al., 2024; Bursztyn et al., 2020)

	any unobserved common cause with the outcome. If these two assumptions are plausible, researchers can use this approach to isolate the unconfounded variation in the explanatory variable. However, the two assumptions cannot be empirically tested.	- Effect of Facebook and Instagram on political beliefs, attitudes, and behavior during the 2020 US election (e.g., Allcott et al., 2024)
Regression-discontinuity design	- The defining feature of the regression discontinuity approach is a running variable, in which there is a sharp change in probability of the explanatory variable of interest being assigned or activated around a threshold value. - This approach has been described as a “locally randomized” experiment (Lee & Lemieux, 2014), because it compares observations around the threshold, motivated by the assumption that these observations are unlikely to systematically differ aside from their status as regards the explanatory variable of interest. Whether this assumption holds will depend on whether the observations can directly manipulate their values on the running variable.	- Effect of Wakefield et al. (1998) on vaccine skepticism (e.g., Motta & Stecula, 2021) - Effect of recession news on consumer behaviors (e.g., Eggers et al., 2021) - Effect of fact checks on Twitter (e.g., Aruguete et al., 2023)
Difference-in-differences and synthetic control	- Difference-in-differences compares differences in outcomes over time between treatment and control groups. With non-random assignment into groups, the analysis relies on the parallel-trends assumption, where the treatment and control groups would have followed the same trend over time in the absence of treatment. - Synthetic controls may be used to better match pre-treatment characteristics, if the parallel-trends assumption is unlikely to hold and no control group is sufficiently similar to the treatment group to act as a counterfactual.	- Demand vs. supply of misinformation on Facebook (Motta et al., 2023) - Effect of misinformation on vaccination and voting behavior (Carrieri et al., 2019) - Effect of fake-news flagging on dissemination behaviours (Ng et al., 2021)

Note. For another introduction and additional references for instrumental variable analysis and regression discontinuity design, see Grosz et al. (2024), and for another introduction and additional references for difference-in-differences and synthetic control, see Rohthbard et al. (2023) and Bonander et al. (2021).

studies from a variety of disciplines. In this way, we are explicitly calling for greater integration of the various misinformation- and conspiracy-adjacent fields, including psychology, economics, political science, data science, sociology, and communications studies.

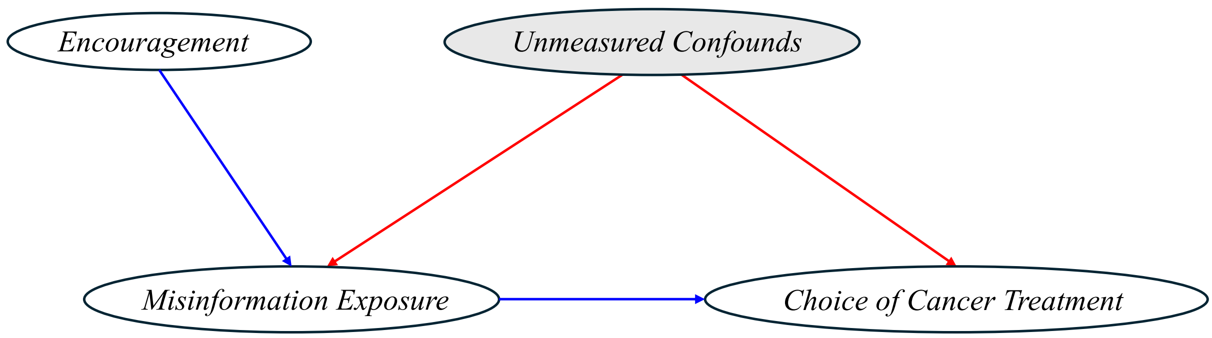
1.2.1 Instrumental-Variable Analysis

To introduce instrumental-variable analysis, consider again the hypothetical research scenario regarding cancer misinformation. It would be unethical to randomly assign vulnerable individuals in an experiment to cancer misinformation, but results from observational studies simply regressing health outcomes on misinformation exposure would be biased by unmeasured confounds (e.g., socioeconomic status may be related to both misinformation exposure and the choice of cancer treatment). If, however, the researchers in this scenario can plausibly justify the existence of a variable that is associated with the outcome of interest (e.g., the choice of cancer treatment) only via its association with the explanatory variable of interest (e.g., exposure to misinformation), instrumental-variable analysis may be considered. Such a variable, for which effects on the outcome can be assumed to be fully mediated by the explanatory variable of interest, would be termed the “instrument” within the context of the analysis. For instance, researchers could randomly encourage a subset of participants to reduce access to relevant sources of misinformation, with the encouragement considered as the instrument (assumed to only affect outcome via its impact on misinformation exposure). See Figure 1 for an illustration.

If the above holds, the data can then be analyzed via two-stage least squares regression. In the first stage, the explanatory variable of interest would be regressed on the instrument; then, in the second stage, the outcome of interest would be regressed on the first-stage predicted values for the explanatory variable of interest. In essence, this approach seeks to overcome unmeasured confounds by exploiting random variation in the hypothesized explanatory variable of interest due to the instrument. This would allow researchers to test for causal effects as applied to the subset of participants that respond to the instrument; formally, this is known as the complier average causal effect. This contrasts with the average causal effect of the encouragement, which would be estimated with an intention-to-treat analysis, the standard in psychology. Critically, instruments

Figure 1

Directed Acyclic Graph Illustrating Instrumental-Variable Analysis



Note. To test for causal effects using instrument-variable analysis, the instrument (e.g., an encouragement) should (1) affect the outcome of interest (e.g., the choice of cancer treatment) only via the explanatory variable of interest (e.g., misinformation exposure) and (2) be unaffected by unmeasured confounding. These two assumptions cannot be directly tested.

need not be experimentally created but can also be identified in the environment. For example, one innovative use of instrumental-variable analysis has been to isolate exogenous (unconfounded) variation in media attention. To illustrate: a significant event A (e.g., a terror attack) will draw substantial media attention, unless another significant concurrent event B, such as a natural disaster, "crowds out" media coverage of event A. In this way, significant event B can serve as the instrument (i.e., the encouragement from the earlier example) to isolate the effect of media attention on significant event A. For example, this approach has allowed researchers to scrutinize if the level of U.S. media coverage of terror events in other countries—which is effectively "manipulated" by the presence versus absence of coverage of a concurrent natural disaster in the U.S. crowding out the terror-event coverage—causally affects the likelihood of future terror attacks in those countries (e.g., Jetter, 2017). In this example, for two-stage least squares, level of U.S. media attention to terrorism during a terrorist event was first regressed on a binary indicator of whether there was a concurrent natural disaster in the U.S., and the estimated value of media attention was then used in the second stage to predict the number of terrorist-attack days in the immediately succeeding week. Nonetheless, regardless of whether instruments are identified or created, researchers must take extra care to consider if core assumptions may have been violated (e.g., whether the instrument may be related to the outcome variable in ways other than the hypothesized explanatory variable of interest; Andrews et al., 2019).

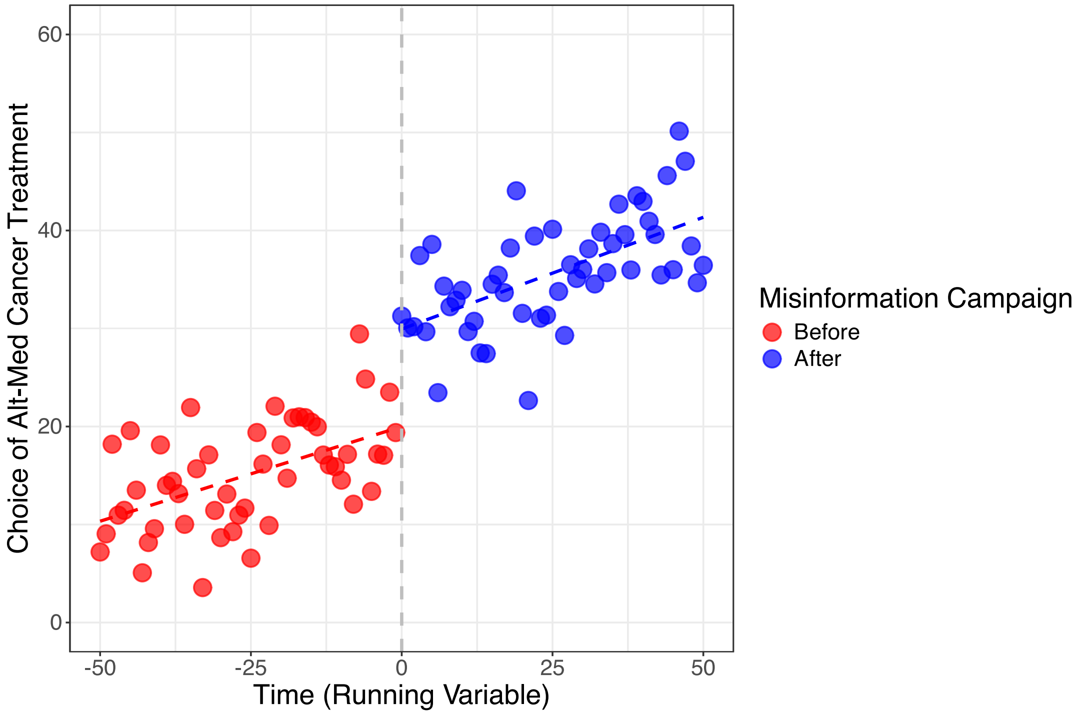
1.2.2 Regression-Discontinuity Design

Another causal-inference approach currently underused in misinformation and conspiracy research is regression-discontinuity design. The defining feature of this approach is a running variable (i.e., a variable whereby there is a sharp increase in probability of the explanatory variable of interest being assigned or activated around a cut-off or threshold value). For example, in the hypothetical study on cancer misinformation, suppose there is a specific date at which a large-scale alternative-medicine misinformation campaign was launched. If researchers seek to study the causal effect of such a misinformation campaign on individuals' choice of cancer treatments, a common observational-study approach would be to ask via questionnaires whether individuals have been exposed to the misinformation campaign and whether they indicate any belief in the relevant false claims. Yet, such a design could again be affected by the presence of unmeasured confounds, as factors such as socio-economic status could impact both access to misinformation as well as the choice of cancer treatments. Instead, relying on a regression-discontinuity design, researchers could make use of time as a running variable, as the probability of an individual being exposed to the misinformation would sharply increase after the launch date of the misinformation campaign. In this case, even if the launch of the misinformation campaign cannot be randomized by researchers, individuals choosing cancer treatments just before versus after the launch of the misinformation campaign are unlikely to systematically differ on other dimensions apart from exposure to the campaign. Thus, researchers can restrict analysis to this subset of individuals, in essence creating a "locally randomized" study (Lee & Lemieux, 2014). See Figure 2 for an illustration.

Using time as a running variable and the publication of the 1998 Wakefield et al. study (which falsely linked the MMR vaccine to autism) as the cut-off, researchers have used a regression-discontinuity design to examine the causal effect of the fraudulent study on subsequent vaccine skepticism (Motta & Stecula, 2021). Importantly, however, the running variable with arbitrary cut-off or threshold values need not be time. For example, regression discontinuity has been applied to study the effects of economic news on consumer behaviors, with the running variable being gross-domestic-product growth,

Figure 2

Hypothetical Data Illustrating Regression-Discontinuity Design



Note. For this hypothetical dataset, each point represents binned observations, the running variable is time (days before and after misinformation campaign), the threshold as marked by the dashed vertical line is therefore at 0, and the outcome of interest is choice of alternative-medicine as cancer treatment. If the assumptions of regression-discontinuity design holds and individuals around the threshold do not systematically differ aside from their probability of being exposure to the misinformation campaign (i.e., a "locally randomized" experiment), the difference in outcomes can be ascribed to the campaign's causal effect.

given that recession announcements are based on the arbitrary cut-off of two consecutive quarters of negative growth (see Eggers et al., 2021). Nonetheless, again, there are assumptions that must be satisfied for the interpretation to be causal. Here, the key assumptions are that the running variable should not be precisely manipulable by the units of analyses (e.g., that individuals cannot decide for themselves whether they are above or below the cut-off), and that the threshold value should not be associated with changes in probability of other relevant variables (e.g., if the probability that an individual is from a higher socio-economic background changes significantly just above or below the ostensibly arbitrary threshold). This is because, in those cases, "local randomization" is no longer a tenable assumption and differences in outcomes can again be due to unmeasured confounds and not the hypothesized causal variable of interest.

1.2.3 Difference-in-Differences and Synthetic Control Designs

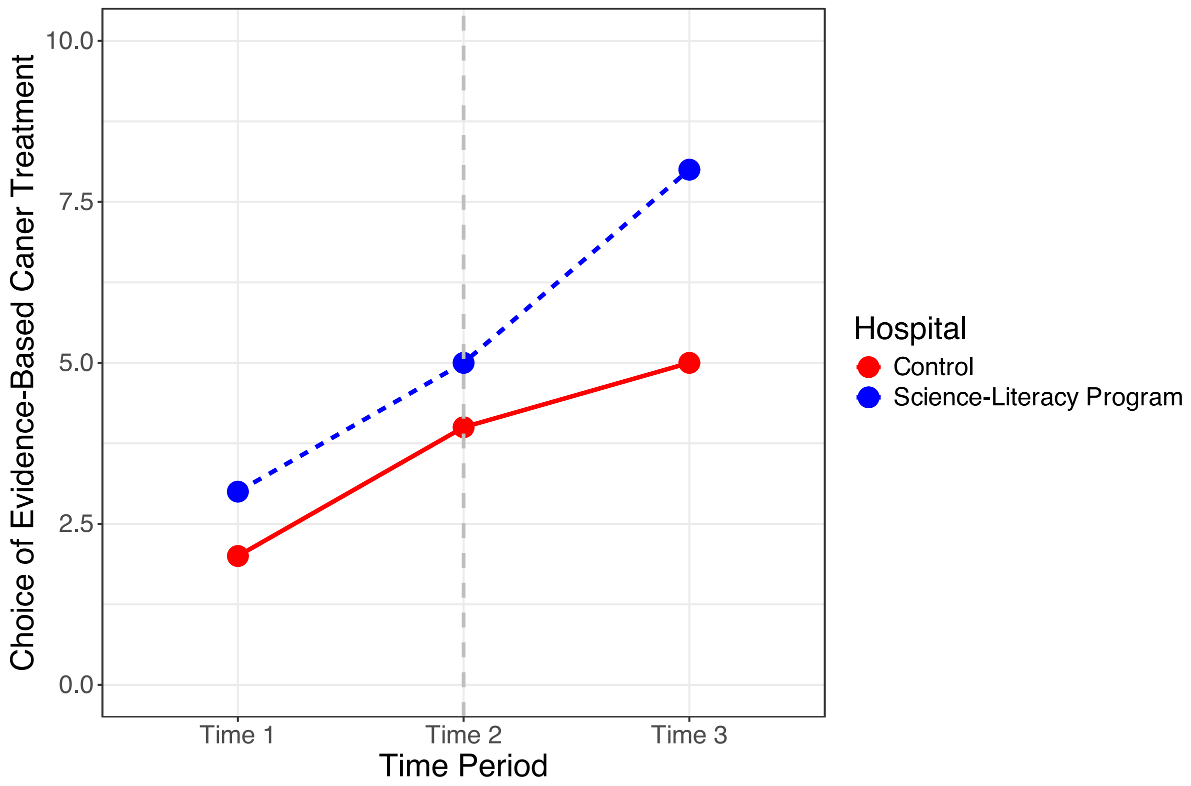
A third approach is the difference-in-differences design. Such a design typically involves the provision of an intervention to certain units of analyses, while others are left out over a period. For example, imagine that two hospitals have decided to implement a science-literacy program to counter the potential influence of cancer misinformation, but one hospital implemented the program earlier than the other. Both hospitals then track patients' science-literacy performance outcomes over time. A

direct comparison of outcomes across the two hospitals would not allow for a causal interpretation, as there may be many dimensions on which the two hospitals differ aside from the implementation of the program. To address this, the difference-in-differences approach aims to exploit the repeated observations over time to provide an estimate of the average causal effect of the intervention, which is defined as the difference between changes in average outcomes over time for the intervention group minus the concurrent changes in the control group acting as a non-randomized counterfactual comparison (i.e., an interaction). In the current example, this would be an increase in the likelihood of an evidence-based treatment being chosen over time in the hospital that implemented the program initially minus any potential (spontaneously occurring) increase in the hospital that at the time did not.

To rule out bias, one key assumption is known as parallel trends, that is, the outcomes for the control group should approximate the path of the intervention group in the absence of the science-literacy program. For example, patients in the two hospitals should be comparable and have similar trajectories prior to

Figure 3

Hypothetical Data Illustrating Difference-in-Differences



Note. For this hypothetical dataset, the key comparison is between the hospital that implemented a science-literacy program at Time 2, as marked by the dashed vertical line, and a control group without such a program. To plausibly ascribe the difference in choice of evidence-based treatment between the two groups across Time 2 and Time 3 to causal effect of the science-literacy program, the parallel-trends assumption should be satisfied. That is, patients' choice of cancer treatment for the two hospitals should follow the same trend, absent the science literacy program. This assumption cannot be tested directed, although researchers can assess trends prior to program implementation (i.e., from Time 1 to Time 2).

implementation of the science-literacy program for one to serve as an adequate counterfactual for the other. In practice, this means that unmeasured confounds are assumed as equal across groups and time. In the long run, many variables can differentially confound the causal effect, so the parallel-trends assumption may only be met within a short timeframe. To test the assumption, one option is to apply an equivalence test to pre-intervention data, although this requires a large sample (e.g., Hartman & Hidalgo, 2018). Other extensions exist and are being actively developed (e.g., Butts & Gardner, 2021; Callaway & Sant'Anna, 2021), such as one that allows for spill-over effects between units (e.g., benefits of one hospital implementing the program affecting other hospitals), or for when the intervention is continuous and varying in intensity (e.g., all hospitals implement science-literacy programs, but the programs differ in terms of the number of lessons). See Figure 3 for an illustration.

Importantly, if no control group is sufficiently similar to the intervention group to act as a counterfactual, a synthetic-control strategy may be a better option. It complements the difference-in-differences approach by introducing an optimally weighted average of a set of potential controls (via minimizing distance functions using pre-intervention covariates, akin to matching algorithms), instead of imposing the parallel-trends assumption (see Abadie et al., 2010). For example, there may simply be no single hospital with patients that have similar trajectories to be used as the control in the difference-in-differences approach, and so an optimally weighted set of hospitals may instead be used. Indeed, in the seminal paper by Abadie and colleagues, a set of 29 states across the U.S. was used to construct a synthetic control that best matched the state of California prior to the introduction of anti-smoking legislation to examine the legislation's causal effects, as no single state was sufficiently similar to California.

Although several studies have used the difference-in-differences approach to study the impact of misinformation and conspiracy theories, as well as associated interventions, to the best of our knowledge only one study to date has attempted to incorporate a synthetic control (Li et al., 2023). In this study, the researchers sought to study the causal effect of Twitter's restrictions on and labelling of some of former U.S. President Trump's tweets during the 2020 U.S. presidential election on subsequent spread of misleading tweets about the election. The synthetic-control strategy was used to construct time-series data of tweets that could be assumed to be maximally similar to the tweets targeted by Twitter (aside from the intervention itself), and the difference in trends between the sets of real and synthetic tweets were then taken as the causal effect of Twitter's intervention. Naturally, as before, whether the causal interpretation is plausible depends on whether the assumptions that underlie the analysis are tenable, which in this case would include adequate similarity of trends across tweet sets (for a recent review, see Abadie et al., 2015).

2. DISCUSSION AND CONCLUDING REMARKS

As may be clear from the preceding section, most research employing the causal-inference approaches covered in this Perspective has been conducted in disciplines such as political science and economics. Speculatively, this is because the subject matter of those disciplines already necessitates more common usage of non-experimental data, leading to greater methodological advancement in causal inference compared to psychology (see Grosz et al., 2020). Nonetheless, as mentioned, randomized experiments remain the "gold standard" for causal inference and decades of psychological research using randomized experiments and observational studies have revealed important insights into a range of phenomena. Naturally, the approaches covered here cannot and should not replace all existing approaches, but they have the potential to act as complements

for researchers to devise additional ways to test and refine psychological theories using real-world data. Indeed, there exists a wide variety of data available from many countries that can be used to diversify our research agenda (e.g., crime and social media data, national surveys). Alongside the broadening of our analytical toolbox, these have the potential to inform both theorizing of naturally occurring behaviors as results of biased or inaccurate media coverage and "fake news", as well as group- and system-level interventions targeting misinformation and conspiracy theories that have so far been neglected in favor of individual-level interventions (e.g., Chater & Loewenstein, 2023).

Nonetheless, while we believe that these causal-inference approaches will be useful tools in the arsenal, we again emphasize that interpretation of results from such analyses as causal depends ultimately on whether the assumptions underlying those analyses are tenable in each particular instance. As such, we also want to draw attention to the potential utility of placebo tests (for sensitivity analysis, see also Oster, 2013). Briefly, placebo tests, within the context of causal inference, refer to analyses in which the primary analysis is replicated but with the units of analysis or outcome measures replaced by alternatives that could not plausibly be affected by the causal variable of interest (Eggers et al., 2021). The goal is thus to assess the credibility of the primary analysis by testing if the strategies employed return placebo estimates that are close to zero or if an effect still emerges. Recent work has additionally proposed the use of pre-registered placebos and equivalence testing that specifies an a-priori range in which differences are deemed inconsequential (Eggers et al., 2023; Hartman & Hidalgo, 2018). Such analyses will be important in theorizing about complex real-world behaviors, where many assumptions that underly causal analysis need to be thoroughly interrogated.

Finally, to conclude, the process of drawing causal inferences, particularly when randomized experiments are not feasible, can be undeniably complex. We hope that the current Perspective contributes a small step towards a more comprehensive understanding of the causes and consequences of misinformation and conspiracy theories in real-world contexts.

3. CONFLICTS OF INTEREST

The authors report no competing interests.

4. AUTHOR CONTRIBUTIONS

L.T. drafted the original manuscript. All authors provided critical revisions.

REFERENCES

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493-505. https://doi.org/10.1198/jasa.2009.ap08746
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science, 59*(2), 495-510. https://doi.org/10.1111/ajps.12116
- Allcott, H., Gentzkow, M., Mason, W., Wilkins, A., Barberá, P., Brown, T., Cisneros, J. C., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gonzáles-Báilon, S., Guess, A. M., Kim, Y., Lazer, D., Malhotra, N., Moehler, D., Nair-Desai, S., El Barj, H., Nyhan, B., ... & Tucker, J. A. (2024). The effects of Facebook and Instagram on the 2020 election: A deactivation experiment. *Proceedings of the National Academy of Sciences, 121*(21), e2321584121. https://doi.org/10.1073/pnas.2321584121
- Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics, 11*, 727-753. https://doi.org/10.1146/annurev-economics-080218-025643
- Aruguete, N., Bachmann, I., Calvo, E.,

Valenzuela, S., & Ventura, T. (2023). Truth be told: How "true" and "false" labels influence user engagement with fact-checks. *New Media & Society.* https://doi.org/10.1177/14614448231193709

- Ash, E., Galletta, S., Hangartner, D., Margalit, Y., & Pinna, M. (2024). The effect of Fox News on health behavior during COVID-19. *Political Analysis, 32*(2), 275-284. https://doi.org/10.1017/pan.2023.21
- Badrinathan, S., & Chauchard, S. (2024). "I Don't Think That's True, Bro!" Social Corrections of Misinformation in India. The International Journal of Press/Politics, 29(2), 394-416. https://doi.org/10.1177/19401612231158770
- Bernheim, B. D., Björkegren, D., Naecker, J., & Pollmann, M. (2022). Causal inference from hypothetical evaluations (No. w29616). National Bureau of Economic Research. https://doi.org/10.3386/w29616
- Bonander, C., Humphreys, D., & Degli Esposti, M. (2021). Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial. *American journal of epidemiology, 190*(12), 2700-2711. https://doi.org/10.1093/aje/kwab211
- Bor, A., & Petersen, M. B. (2022). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American political science review, 116*(1), 1-18. https://doi.org/10.1017/S0003055421000885
- Boulianne, S., & Humprecht, E. (2023). Perceived exposure to misinformation and trust in institutions in four countries before and during a pandemic. *International Journal of Communication, 17,* 24. https://doi.org/1932–8036/20230005
- Bursztyn, L., Rao, A., Roth, C. P., & Yanagizawa-Drott, D. H. (2020). Misinformation during a pandemic (No. w27417). *National Bureau of Economic Research.* https://doi.org/10.3386/w27417
- Butts, K., & Gardner, J. (2021). {did2s}: Two-stage difference-in-differences. *arXiv.* https://doi.org/10.48550/arXiv.2109.05913
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001
- Carrieri, V., Madio, L., & Principe, F. (2019). Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics, 28*(11), 1377-1382. https://doi.org/10.1002/hec.3937
- Chater, N., & Loewenstein, G. (2023). Where next for behavioral public policy?. *Behavioral & Brain Sciences, 46.* https://doi.org/10.1017/S0140525X23002091
- Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology, 40,* 3-35. https://doi.org/10.1111/pops.12568
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*(1), 13-29. https://doi.org/10.1038/s44159-021-00006-y
- Ecker, U. K. H., Tay, L. Q., Roozenbeek, J., van der Linden, S., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). *Why misinformation must not be ignored.* https://osf.io/8a6cj/download
- Eggers, A. C., Ellison, M., & Lee, S. S. (2021). The economic impact of recession announcements. *Journal of Monetary Economics, 120,* 40-52. https://doi.org/10.1016/j.jmoneco.2021.03.002
- Eggers, A. C., Tuñón, G., & Dafoe, A. (2023). Placebo tests for causal inference. *American Journal of Political Science.* https://doi.org/10.1111/ajps.12818
- Enders, A. M., Uscinski, J. E., Seelig, M. I., Klofstad, C. A., Wuchty, S., Funchion, J. R., Murthi, M. N., Premaratne, K., & Stoler, J. (2021). The relationship between social media use and beliefs in conspiracy theories and misinformation. *Political Behaviour,* 1-24. https://doi.org/10.1007/s11109-021-09734-6

- Goreis, A., & Voracek, M. (2019). A systematic review and meta-analysis of psychological research on conspiracy beliefs: Field characteristics, measurement instruments, and associations with personality traits. *Frontiers in Psychology, 10*, 425400. https://doi.org/10.3389/fpsyg.2019.00205
- Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Willmes, A., & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science, 7*(1), 25152459231218610. https://doi.org/10.1177/25152459231218610
- Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science, 15*(5), 1243-1255. https://doi.org/10.1177/1745691620921521
- Hartman, E., & Hidalgo, F. D. (2018). An equivalence approach to balance and placebo tests. *American Journal of Political Science, 62*(4), 1000-1013. https://doi.org/10.1111/ajps.12387
- Haslam, S. A., Reicher, S. D., & Platow, M. J. (2020). *The new psychology of leadership: Identity, influence and power.* Routledge.
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751
- Jetter, M. (2017). The effect of media attention on terrorism. *Journal of Public Economics, 153*, 32-48. https://doi.org/10.1016/j.jpubeco.2017.07.008
- Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media+ Society, 7*(1), 2056305121988928. https://doi.org/10.1177/2056305121988928
- Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences, 120*(7), e2210666120. https://doi.org/10.1073/pnas.2210666120
- Lee, D. S., & Lemieux, T. (2014). Regression discontinuity designs in social sciences. *The SAGE Handbook of Regression Analysis and Causal Inference*, 301-27.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008
- Li, Z., Cao, J., Adams-Cohen, N., Alvarez, R.M. (2023). The Effect of Misinformation Intervention: Evidence from Trump’s Tweets and the 2020 Election. In: Ceolin, D., Caselli, T., Tulin, M. (eds) *Disinformation in Open Online Media. MISDOOM 2023*. Lecture Notes in Computer Science, vol 14397. Springer, Cham. https://doi.org/10.1007/978-3-031-47896-3_7
- Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science, 16*(4), 854-863. https://doi.org/10.1177/1745691620974773
- Loomba, S., De Figueiredo, A., Piatek, S. J., De Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature human behaviour*, *5*(3), 337-348. https://doi.org/10.1038/s41562-021-01056-1
- Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review, 86*(3), 532-565. https://doi.org/10.1177/00031224211004187
- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95.

https://psycnet.apa.org/doi/10.1037/h0056029

- Marinescu, I. E., Lawlor, P. N., & Kording, K. P. (2018). Quasi-experimental causality in neuroscience and behavioural research. *Nature human behaviour*, *2*(12), 891-898. https://doi.org/10.1038/s41562-018-0466-5
- Motta, M., Hwang, J., & Stecula, D. (2023). What goes down must come up? Pandemic-related misinformation search behavior during an unplanned Facebook outage. *Health Communication*, 1-12. https://doi.org/10.1080/10410236.2023.2254583
- Motta, M., & Stecula, D. (2021). Quantifying the effect of Wakefield et al.(1998) on skepticism about MMR vaccine safety in the US. *PLOS ONE*, *16*(8), e0256395. https://doi.org/10.1371/journal.pone.0256395
- Murphy, G., de Saint Laurent, C., Reynolds, M., Aftab, O., Hegarty, K., Sun, Y., & Greene, C. M. (2023). What do we study when we study misinformation? A scoping review of experimental research (2016-2022). *Harvard Kennedy School Misinformation Review.* https://doi.org/10.37016/mr-2020-130
- Murphy, J. J., Stevens, T. H., & Yadav, L. (2010). A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. *Environmental and Resource Economics*, *47*, 111-123. https://doi.org/10.1007/s10640-010-9367-4
- Newman, E. J., Swire-Thompson, B., & Ecker, U. K. H. (2022). Misinformation and the sins of memory: False-belief formation and limits on belief revision. https://doi.org/10.1037/mac0000090
- Ng, K. C., Tang, J., & Lee, D. (2021). The effect of platform intervention policies on fake news dissemination and survival: An empirical examination. *Journal of Management Information Systems*, *38*(4), 898-930. https://doi.org/10.1080/07421222.2021.1990612
- Oster, E. (2013). *Unobservable selection and coefficient stability: Theory and validation* (No. w19054). National Bureau of Economic Research. https://ssrn.com/abstract=2266720
- Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), 2333. https://doi.org/10.1038/s41467-022-30073-5
- Pillai, R. M., & Fazio, L. K. (2021). The effects of repeating false and misleading information on belief. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(6), e1573. https://doi.org/10.1002/wcs.1573
- Prike, T., Butler, L. H., & Ecker, U. K. H. (2024). Source-credibility information and social norms improve truth discernment and reduce engagement with misinformation online. *Scientific Reports*, *14*(1), 6900. https://doi.org/10.1038/s41598-024-57560-7
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27-42. https://doi.org/10.1177/2515245917745629
- Rothbard, S., Etheridge, J. C., & Murray, E. J. (2023). A tutorial on applying the difference-in-differences method to health data. *Current Epidemiology Reports*, 1-11. https://doi.org/10.1007/s40471-023-00327-x
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688. https://psycnet.apa.org/doi/10.1037/h0037350
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences, 34*, 81-87. https://doi.org/10.1016/j.cobeha.2020.01.003
- Tay, L. Q., Lewandowsky, S., Hurlstone, M. J.,

Kurz, T., & Ecker, U. K. H. (2023). A focus shift in the evaluation of misinformation interventions. *Harvard Kennedy School Misinformation Review.* https://doi.org/10.37016/mr-2020-124

- Tay, L. Q., Lewandowsky, S., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2024). Thinking clearly about misinformation. *Communications Psychology*, *2*(1), 4. https://doi.org/10.1038/s44271-023-00054-5
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008